Genomics Lab

In class, we have talked multiple times about how genetic information has enriched herpetology. Recent advances in genetic sequencing technology have made a new type of information, whole genome sequences, increasingly available. This information allows us to push the boundaries of the field. Today, you will be exploring complete genome sequences from two species of herps: the green anole (*Anolis carolinensis*) and the African clawed frog (*Xenopus tropicalis*). The goal of this lab is to experience how whole genome sequences can be used as tools to learn about the biology of herps.

Figure 1. *Anolis carolinensis*

Figure 2. *Xenopus tropicalis*

The two species in this lab were the two first herp genomes to be completed. To carry out this lab, you will access genetic information for these two (and other) species from two sources: the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), which is geared towards viewing complete genome sequences and doing simple analyses on them, and GenBank (http://www.ncbi.nlm.nih.gov/), a database for all types of genetic sequence data for a wide range of species. You will carry out a few activities that will allow you to see the genome sequences that are publicly available, and carry out some simple analyses on these datasets.

Part 1. *Xenopus tropicalis* (western clawed frog)

Figure 3. Ovaries and oocytes for *X. laevis*. Bottom oocyte is ~1 mm in diameter

A. Background

Xenopus frogs are widely used in biomedical research. *Xenopus* oocytes are much larger than those of other organisms, like humans, and their embryos are large and easy to manipulate. This has made *Xenopus* a model system for two areas of biological research. First, these frogs are widely used in cell biology; much of what we know about the processes of mitosis and meiosis comes from researchers working on *Xenopus*. Second, these frogs are common models for developmental biology. Both of these areas of research usually use *X. laevis*, the African clawed frog. However, *X. laevis* is a tetraploid species with a large genome: 4N = 36, 3.1 gigabases (GB), where one GB = one billion (1×10^9) base pairs of DNA. This is about three times the size of the amphibian with the smallest genome (that we know), *Limnodynastes ornatus*, which has a genome size of about 1 GB. By comparison, the human genome is about 3.4 GB.

The large size of the *X. laevis* genome, as well as the fact that this frog is tetraploid, makes sequencing its whole genome difficult. Most other species in *Xenopus* are polyploid, with many chromosomes and a very large genome. One species, *Xenopus ruwenzoriensis*, has 108 chromosomes, and may be hexaploid (6N = 108), although some of the sextets of chromosomes are heterogeneous. However, one species is diploid with a relatively small genome: *Xenopus tropicalis,* the western clawed frog (2N = 20, 1.7 GB). There is some taxonomic controversy within *Xenopus*, and some taxonomists place *tropicalis* in a separate genus, *Silurana*. I will use the name *Xenopus* here to be concordant with the genome sequencing sites. This species is found in forests of west Africa. This is the species whose whole genome has been sequenced.

A.Biology of Xenopus tropicalis

We know less about the biology of this species in the wild than we do about it's genetics and cell biology. You can find a good summary of the available information on AmphibiaWeb (http://amphibiaweb.org). Find the AmphibiaWeb entry for this species, and use it to answer the questions below.

QUESTION 1: Natural history of *Xenopus*

a. What is the taxonomy of this frog species? List some facts about the family to which this species belongs.

b. What are the phylogenetic relationships between this family of frogs and other frogs? HINT: if you don't remember, refer to my notes from class or look at the tree of life website.

c. What is the distribution of this species?

C. Genome of Xenopus tropicalis

You can find a general description of the *X. tropicalis* genome at the Joint Genome Institute's (JGI) website (http://genome.jgi-psf.org/Xentr4/Xentr4.home.html). For now, all that you really need to know are two crucial pieces of information. First, this species is diploid, with 2N = 20, and has a genome of approximately 1.7 GB. Second, the genome sequence for this species has been completely sequenced, but not completely assembled into the ten unique chromosomes characteristic of the species. Instead, the genome sequence is summarized in a series of "scaffolds", long pieces of DNA that haven't yet been gathered into chromosomes. The database currently contains 19,501 scaffolds, partially overlapping; most of the genome is covered by scaffolds 1-272, all of which are long (each at least 1.5 megabases, Mb, where 1 Mb = 1 x 10^6 base pairs).

The easiest way to view the genome of this animal is through the UCSC Genome Browser. Go to the main page for browsing genomes through this resource (http://genome.ucsc.edu/cgi-bin/hgGateway). You will see a place to choose the part of the genome to look at near the top of the screen. Enter the information below to look at scaffold 1 of the *X. tropicalis* genome.

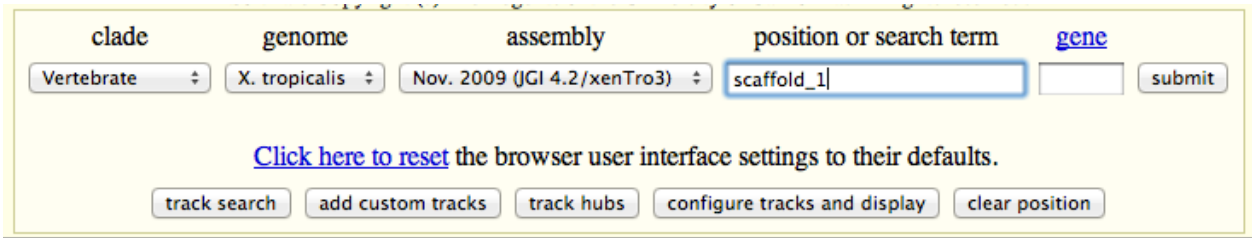

Figure 4. How to get to scaffold 1 of *X. tropicalis.*

You will now see a display of scaffold one on your screen. You can click on various parts of this display to get more information; briefly, the screen shows places in this scaffold that look like genes, and which genes they might be. Use the information from this screen to answer the questions below.

QUESTION 2: Part of the *Xenopus* genome: scaffold 1 (2 points)
a. Zoom in until you get to the level of individual base pairs, then out again to see the whole scaffold. How long is this section of the *Xenopus* genome? What percentage of the entire genome does this represent?

b. Below the long section of "contig yes" blocks (don't worry about those), there are genes marked using colored blocks. List ten genes that are found on scaffold 1.
For one of these genes, try to find some more information about it (e.g. what it does in other species, perhaps). Summarize the results of your search.

c. At the very bottom of the figure you will see a plot summarizing how conserved (similar) this stretch of frog DNA is to the DNA of other species. Do any of these organisms seem more similar to the frog than any other? Does this make sense from a phylogenetic perspective?

D. Finding known genes in Xenopus tropicalis

You will now look for a known gene in the *Xenopus* genome. In the **position or search term** box, type “hemoglobin alpha 1.” This is the alpha subunit of hemoglobin, a protein that serves the same function in frogs as it does in humans. You will see a list of hits from this search; pick the one that is called hba1 in the RefSeq list.

Have a look at this gene. Feel free to scroll around and zoom in or out.

QUESTION 3:
a. How long is this gene?

b. The filled boxes are the exons of the gene, while the dashed lines represent introns. Estimate the total length of the exons in this gene.

Now, let’s make sure this sequence matches the gene that we think it does. First, go to the correct zoom level so that you span the original length of this gene - the easiest way to do that, if you’ve been messing with the zoom and scrolling, is to run your search again. Now click on the “DNA” button, in the blue bar near the top of the window, and click “get DNA on the next page. This will give you the genetic sequence for hba1 in *X. tropicalis*. Keep this browser window open so that you can access this sequence in a couple minutes.

Now, open a new browser window and go to GenBank (http://www.ncbi.nlm.nih.gov/). Once you are there, click on “blast” in the blue banner at the top of the screen. Blast is a way to find sequences that are close to a particular genetic sequence of interest. Next,. choose the program called **nucleotide_blast**. This will bring up a window where you can enter your data.

Go back to your hba1 sequence, highlight the whole sequence and associated text at the beginning of the screen, and paste this sequence into the large box at the top of the blast search page. Change the database to “others,” and choose nucleotide collection (nr/nt). This will search all of genbank for sequences similar to your frog sequence. In the bottom program selectio box, choose “Somewhat similar sequences (blastn).” Finally, push the big blue “BLAST button at the bottom of the screen. Now use your results to answer the next set of questions.

QUESTION 4:
a. Look at the results of your blast search. What organisms have similar sequences in their genomes? What genes are represented? Interpret this data in terms of relationships among species and genes.

b. Click on “Distance tree of results.” This will show the sequences found by blast on a rough phylogenetic tree. Which sequences are most similar to the Xenopus sequence? Is this phylogeny the same as the one presented in class? What might have caused the differences?

E. Identify an unknown gene in Xenopus tropicalis

Next, I will send you to a particular location in the frog genome, and you will use BLAST to identify a “mystery gene.”

Go to the following location in the *X. tropicalis* genome:
scaffold_29:4,305,430-4,308,848
Use the DNA button to obtain a sequence for this gene, and then run a blast search using this sequence. Use the results of the blast search to the questions below. It might be helpful to redo your BLAST search and limit it to the human genome.

QUESTION 5: What is this gene? What is its importance in humans? Do you think this has any implications for the scope and relevance of *Xenopus* genomic research?

Part 2. *Anolis carolinensis* (green anole)

We've talked in class about *Anolis* lizards as a model system in ecology and evolution. Species of anoles are also used in a wide range of studies, from physiology to neurobiology to macroevolution. Thus, the idea to sequence the whole genome of the green anole (*Anolis carolinensis*) was initiated a couple of years ago. Similar to the frog, we have a series of scaffolds for the anole that are not combined into chromosomes.

A. Do anoles have a MC1R gene?

Using the search box (Figure 4), search the anole genome for a MC1R gene. This is a gene that is crucial for coloration in a wide range of animals. Look over the results from this search, and use them to answer this question.

QUESTION 6: We talked about MC1R during the first lecture of class. Is MC1R present in Anolis? What other species have similar MC1R sequences?

B. Identifying an unknown gene in *Anolis*.

Go to this location in the anole genome: scaffold_580:774,737-781,471. Answer the question below. (HINT: you might want to run a BLAST search on this sequence).

QUESTION 7: What gene is present in the genome sequence of *A. carolinensis* in this region? What is the function of this gene?

Part 3: Future directions in genomics

Genomes are increasingly cheap to obtain, and represent a major realm for future research in herpetology. Look at the list of available genomes in the UCSC genome browser.

QUESTION 8: What genome should we sequence next? Why?